2/5/1
DIALOG(R)File 351:Derwent

011076976 **Image available**
WPI Acc No: 1997-054900/199706
XRPX Acc No: N97-044988

Speech classifying method esp. method of encoding speech signals - involving forming sub frames and dividing each into one of several typical classes for speech encoding

Patent Assignee: DEUT TELEKOM AG (DEBP)
Inventor: STEGMANN J

Number of Countries: 021 Number of Patents: 005
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
EP 751495	A2	19970102	EP 96104213	A	19960316	199706	B
DE 19538852	A1	19970102	DE 1038852	A	19951019	199706	
NO 9601636	A	19970102	NO 961636	A	19960424	199711	
CA 2188369	A	19970420	CA 2188369	A	19961021	199734	
US 5781881	A	19980714	US 96734657	A	19961021	199835	

Priority Applications (No Type Date): DE 1038852 A 19951019; DE 1023598 A 19950630
Cited Patents: No-SR.Pub
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
EP 751495	A2	G	8	G10L-009/16	
Designated States (Regional): AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE					
DE 19538852	A1		8	G10L-007/02	
NO 9601636	A			G10L-007/02	
CA 2188369	A			G10L-009/14	
US 5781881	A			G10L-007/02	

Abstract (Basic): EP 751495 A

The method classifies speech, in particular speech signals, for the adaptive control of a speech encoding process. This encoding reduces the bit rate while keeping the speech quality the same, or increases the quality while keeping the bit rate the same. After segmenting the speech signal for each frame, a wavelet transformation is calculated. Using adaptive thresholds, a set of parameters is derived which control a state model. The speech frames are divided into sub-frames. Each sub-frame is divided into one of several typical classes for the speech encoding.

The speech signal may be divided into segments of constant length. To reduce the edge effects with the wavelet transformation, either the segment at the boundaries is reflected or the wavelet transformation is calculated at smaller intervals. The frames are preferably shifted such that the segments overlap, or at the edges the segments are filled with previous or predicted sample values.

ADVANTAGE - Is less sensitive to background noise and has reduced complexity. Provides high resolution output and high speech quality.

Dwg.1/2

Title Terms: SPEECH; CLASSIFY; METHOD; METHOD; ENCODE; SPEECH; SIGNAL; FORMING; SUB; FRAME; DIVIDE; ONE; TYPICAL; CLASS; SPEECH; ENCODE
Derwent Class: P86; W02; W04
International Patent Class (Main): G10L-007/02; G10L-009/14; G10L-009/16
International Patent Class (Additional): H03M-007/30
File Segment: EPI; EngPI

für A'im 98P1695

(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) **Offenlegungsschrift**

(10) **DE 195 38 852 A 1**

B1

(51) Int. Cl.[6]:
G 10 L 7/02
H 03 M 7/30

(21) Aktenzeichen: 195 38 852.6
(22) Anmeldetag: 19. 10. 95
(43) Offenlegungstag: 2. 1. 97

DE 195 38 852 A1

(30) Innere Priorität: (32) (33) (31)
30.06.95 DE 195235983

(71) Anmelder:
Deutsche Telekom AG, 53113 Bonn, DE

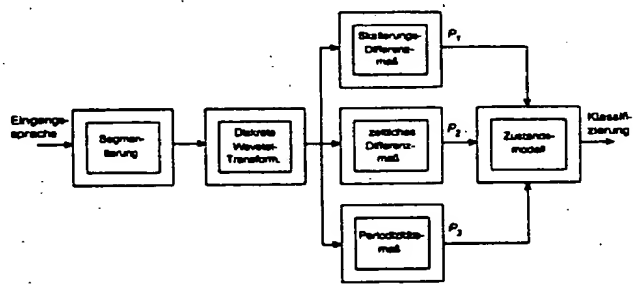
(72) Erfinder:
Stegmann, Joachim, Dipl.-Ing., 64283 Darmstadt, DE

(56) Für die Beurteilung der Patentfähigkeit in Betracht zu ziehende Druckschriften:

DE	195 05 435 C1
DE	44 27 656 C1
DE	44 40 838 A1
DE	44 37 790 A1
DE	43 40 591 A1
DE	43 15 315 A1
DE	43 15 313 A1
DE	42 37 563 A1
DE	42 03 436 A1
GB	22 72 554 A
EP	05 19 802 A1

(54) Verfahren und Anordnung zur Klassifizierung von Sprachsignalen

(57) Es wird ein Verfahren und eine Anordnung zur Klassifizierung von Sprache auf Basis der Wavelet-Transformation für niederratige Sprachcodierverfahren beschrieben. Das Verfahren bzw. die Anordnung als robuster Klassifizierer von Sprachsignalen für die signalangepaßte Steuerung von Sprachcodierverfahren zur Senkung der Bitrate bei gleichbleibender Sprachqualität oder zur Erhöhung der Qualität bei gleicher Bitrate ist dadurch charakterisiert, daß nach Segmentierung des Sprachsignals für jeden Rahmen eine Wavelet-Transformation berechnet wird, aus der mit Hilfe adaptiver Schwellen ein Satz Parameter ermittelt wird, die ein Zustandsmodell steuern, das den Rahmen in gegebenenfalls kürzere Unterrahmen aufteilt und jeden dieser Unterrahmen in eine von mehreren, für die Sprachcodierung typische Klassen einteilt. Das Sprachsignal wird auf Basis der Wavelet-Transformation für jeden Zeitrahmen klassifiziert. Dadurch kann sowohl eine hohe Auflösung im Zeitbereich (Lokalisierung von Pulsen) als auch im Frequenzbereich (gute Mittelwerte) erreicht werden. Dieses Verfahren und der Klassifizierer eignen sich deshalb besonders zur Steuerung bzw. Auswahl von Codebüchern in einem niederratigen Sprachcoder. Sie weisen überdies eine hohe Unempfindlichkeit gegenüber Hintergrundgeräuschen sowie eine niedrige Komplexität auf.



DE 195 38 852 A1

Beschreibung

Die Erfindung betrifft ein Verfahren zur Klassifizierung von Sprachsignalen nach dem Oberbegriff des Patentanspruchs 1 sowie eine Schaltungsanordnung zur Durchführung des Verfahrens.

Sprachcodierverfahren und zugehörige Schaltungsanordnungen zur Klassifizierung von Sprachsignalen für Bitraten unterhalb von 8 kbit pro Sekunde gewinnen zunehmend an Bedeutung.

Die Hauptanwendungen hierfür sind unter anderem bei Multiplexübertragung für bestehende Festnetze und in Mobilfunksystemen der dritten Generation zu sehen. Auch für die Bereitstellung von Diensten wie zum Beispiel Videophonie werden Sprachcodierverfahren in diesem Datenratenbereich benötigt.

Die meisten derzeit bekannten, hochqualitativen Sprachcodierverfahren für Datenraten zwischen 4 kbit/s und 8 kbit/s arbeiten nach dem Prinzip des Code Excited Linear Prediction (CELP)-Verfahrens wie es von Schroeder, M.R., Atal, B.S.: Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates, in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1985, erstmals beschrieben worden ist. Dabei wird das Sprachsignal durch lineare Filterung von Anregungsvektoren aus einem oder mehreren Codebüchern synthetisiert. In einem ersten Schritt werden die Koeffizienten des Kurzzeit-Synthesefilters durch LPC-Analyse aus dem Eingangs-Sprachvektor ermittelt und dann quantisiert. Im Anschluß daran werden die Anregungscodebücher durchsucht, wobei als Optimierungskriterium der perzeptuell gewichtete Fehler zwischen Original- und synthetisiertem Sprachvektor verwendet wird (⇒ Analyse durch Synthese). Übertragen werden schließlich nur die Indizes der optimalen Vektoren, aus denen der Decoder den synthetisierten Sprachvektor wieder erzeugen kann.

Viele dieser Codierverfahren, wie zum Beispiel der neue 8 kbit/s Sprachcoder von ITU-T, beschrieben in der Literaturstelle Study Group 15 Contribution — Q. 12/15: Draft Recommendation G.729 — Coding Of Speech at 8 kbit/s using Conjugate-Structure-Algebraic-Code-Excited-Linear-Predictive (CS-ACELP) Coding, 1995, arbeiten mit einer festen Kombination von Codebüchern. Diese starre Anordnung berücksichtigt nicht die starken zeitlichen Änderungen der Eigenschaften des Sprachsignals und benötigt zur Codierung im Durchschnitt mehr Bits als erforderlich. Zum Beispiel bleibt das nur zur Codierung von periodischen Sprachabschnitten erforderliche adaptive Codebuch auch während eindeutig nichtperiodischer Segmente eingeschaltet.

Um zu niedrigeren Datenraten im Bereich um 4 kbit/s bei möglichst wenig abfallender Qualität zu gelangen, wurde deshalb in anderen Veröffentlichungen, zum Beispiel in Wang, S., Gersho, A.: Phonetically-Based Vector Excitation Coding of Speech at 3.6 kbit/s, Proceedings of IEEE International Conference On Acoustics, Speech and Signal Processing, 1989, vorgeschlagen, das Sprachsignal vor der Codierung in verschiedene typische Klassen einzuordnen. Im Vorschlag für das GSM-Halbratensystem wird das Signal auf Basis des Langzeit-Prädiktionsgewinns rahmenweise (alle 20 ms) in stimmhafte und stimmlose Abschnitte mit jeweils angepaßten Codebüchern eingeteilt, wodurch die Datenrate für die Anregung gesenkt und die Qualität gegenüber dem Vollratensystem weitgehend gleich bleibt. Bei einer allgemeineren Untersuchung wurde das Signal in die Klassen stimmhaft, stimmlos und Onset eingeteilt. Dabei wurde die Entscheidung rahmenweise (hier 11,25 ms) auf Basis von Parametern — wie unter anderem Nulldurchgangsrate, Reflexionskoeffizienten, Energie — durch lineare Diskriminierung gewonnen, siehe zum Beispiel Campbell, J., Tremain, T.: Voiced/Unvoiced Classification Of Speech with Application to the U.S. Gouvernment LPC-10e Algorithm, Proceedings of IEEE International Conference On Acoustics, Speech and Signal Processing, 1986. Jeder Klasse wird wiederum eine bestimmte Kombination von Codebüchern zugeordnet, so daß die Datenrate auf 3,6 kbit/s bei mittlerer Qualität gesenkt werden kann.

All diese bekannten Verfahren ermitteln das Ergebnis ihrer Klassifizierung aus Parametern, die durch Berechnung von Zeitmittelwerten aus einem Fenster konstanter Länge gewonnen wurden. Die zeitliche Auflösung ist also durch die Wahl dieser Fensterlänge fest vorgegeben. Verringert man die Fensterlänge, so sinkt auch die Genauigkeit der Mittelwerte. Erhöht man dagegen die Fensterlänge, so kann der zeitliche Verlauf der Mittelwerte dem Verlauf des instationären Sprachsignals nicht mehr folgen. Dies gilt besonders für stark instationäre Übergänge (Onsets) von stimmlosen auf stimmhafte Sprachabschnitte. Gerade die zeitlich richtige Reproduktion der Lage der ersten signifikanten Pulse stimmhafter Abschnitte ist aber wichtig für die subjektive Beurteilung eines Codierverfahrens. Weitere Nachteile herkömmlicher Klassifizierungsverfahren sind oftmals eine hohe Komplexität oder starke Abhängigkeit von in der Praxis immer vorhandenen Hintergrundgeräuschen.

Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren und einen Klassifizierer von Sprachsignalen für die signalangepaßte Steuerung von Sprachcodierverfahren zur Senkung der Bitrate bei gleichbleibender Sprachqualität bzw. zur Erhöhung der Qualität bei gleicher Bitrate zu schaffen, die das Sprachsignal mit Hilfe der Wavelet-Transformation für jeden Zeitraum klassifizieren, wobei sowohl eine hohe Auflösung im Zeitbereich als auch im Frequenzbereich erreicht werden soll.

Die Lösung für das erfindungsgemäße Verfahren ist im Kennzeichen des Patentanspruchs 1 charakterisiert und die für den Klassifizierer im Kennzeichen des Patentanspruchs 5.

Weitere Lösungen bzw. Ausgestaltungen der Erfindung ergeben sich aus den Kennzeichen der Patentansprüche 2—4.

Hier werden ein Verfahren und eine Anordnung beschrieben, die das Sprachsignal auf Basis der Wavelet-Transformation für jeden Zeitrahmen klassifizieren. Dadurch kann — den Anforderungen des Sprachsignals entsprechend — sowohl eine hohe Auflösung im Zeitbereich (Lokalisierung von Pulsen) als auch im Frequenzbereich (gute Mittelwerte) erreicht werden. Die Klassifizierung eignet sich deshalb besonders zur Steuerung bzw. Auswahl von Codebüchern in einem niederratigen Sprachcoder. Dabei weist das Verfahren und die Anordnung eine hohe Unempfindlichkeit gegenüber Hintergrundgeräuschen sowie eine niedrige Komplexität auf. Bei der Wavelet-Transformation handelt es sich — ähnlich der Fourier-Transformation — um ein mathematisches Verfahren zur Bildung eines Modells für ein Signal oder System. Im Gegensatz zur Fourier-Transformation kann man aber im Zeit- und Frequenz- bzw. Skalierungsbereich die Auflösung den Anforderungen entsprechend flexibel anpassen. Die Basisfunktionen der Wavelet-Transformation werden durch Skalierung und

Verschiebung aus einem sogenannten Mother-Wavelet erzeugt und haben Bandpaßcharakter. Die Wavelet-Transformation ist somit erst durch Angabe des zugehörigen Mother-Wavelets eindeutig definiert. Hintergründe und Details zur mathematischen Theorie sind beispielsweise aufgezeigt von Rioul O., Vetterli, M.: Wavelets and Signal Processing, IEEE Signal Processing Magazine, Oct. 1991.

Aufgrund ihrer Eigenschaften eignet sich die Wavelet-Transformation gut zur Analyse instationärer Signale. Ein weiterer Vorteil ist die Existenz schneller Algorithmen, mit denen eine effiziente Berechnung der Wavelet-Transformation durchgeführt werden kann. Erfolgreiche Anwendungen im Bereich der Signalverarbeitung findet man unter anderem in der Bildcodierung, bei Breitbandkorrelationsverfahren (zum Beispiel für Radar) sowie zur Sprachgrundfrequenzschätzung, wie unter anderem aus den folgenden Literaturstellen hervorgeht. Mallat, S., Zhong, S.: Characterization of Signals from Multiscale Edges, IEEE Transactions on Pattern Analysis and Machine Intelligence, July, 1992 sowie Kadambe, S. Boudreaux-Bartels, G.F.: Applications of the Wavelet Transform for Pitch Detection of Speech Signals, IEEE Transactions on Information Theory, March 1992.

Die Erfindung wird im folgenden anhand eines Ausführungsbeispiels näher beschrieben. Für die Beschreibung des Verfahrens soll der prinzipielle Aufbau eines Klassifizierers nach Fig. 1 verwendet werden. Zunächst erfolgt die Segmentierung des Sprachsignals. Das Sprachsignal wird in Segmente konstanter Länge eingeteilt, wobei die Länge der Segmente zwischen 5 ms und 40 ms betragen soll. Zur Vermeidung von Randeffekten bei der sich anschließenden Transformation kann eine der drei folgenden Techniken angewandt werden:

— Das Segment wird an den Grenzen gespiegelt.
— Die Wavelet-Transformation wird im kleineren Intervall (L/2,N−L/2) berechnet und der Rahmen nur um den konstanten Versatz L/2 verschoben, so daß die Segmente überlappen. Dabei ist L die Länge eines auf den zeitlichen Ursprung zentrierten Wavelets, wobei die Bedingung $N > L$ gelten muß.
— An den Rändern des Segmentes wird mit den vorangegangenen bzw. zukünftigen Abtastwerten aufgefüllt.

Danach erfolgt eine diskrete Wavelet-Transformation. Für ein solches Segment s(k), wird eine zeitdiskrete Wavelet-Transformation (DWT) Sh(m, n) bezüglich eines Wavelets h(k) mit den ganzzahligen Parametern Skalierung in und Zeitverschiebung n berechnet. Diese Transformation ist durch

$$S_h(m,n) = \sum_{k=N_u}^{N_0} s(k)\, h^*\left(\frac{k - n a_0^m}{a_0^m}\right)$$

definiert, wobei N_u und N_0 die durch die gewählte Segmentierung vorgegebene untere bzw. obere Grenze des Zeitindex k darstellen. Die Transformation muß nur für den Skalierungsbereich $0 < m < M$ und den Zeitbereich im Intervall (O, N) berechnet werden, wobei die Konstante M in Abhängigkeit von a_0 so groß gewählt werden muß, daß die niedrigsten Signalfrequenzen im Transformationsbereich noch ausreichend gut repräsentiert werden.

Zur Klassifizierung von Sprachsignalen reicht es in der Regel aus, das Signal zu dyadischen Skalierungen ($a_0 = 2$) zu betrachten. Läßt sich das Wavelet h(k) durch eine sogenannte "Multiresolution-Analyse" gemäß Rioul, Vetterli mittels einer iterierten Filterbank darstellen, so kann man zur Berechnung der dyadischen Wavelet-Transformation in der Literatur angegebene effiziente, rekursive Algorithmen verwenden. In diesem Fall ($a_0 = 2$) ist eine Zerlegung bis maximal $M = 6$ ausreichend. Für die Klassifizierung eignen sich besonders Wavelets mit wenigen signifikanten Oszillationszyklen, aber dennoch möglichst glattem Funktionsverlauf. Beispielsweise können kubische Spline-Wavelets oder orthogonale Daubechies-Wavelets geringer Länge verwendet werden.

Hiernach erfolgt die Klasseneinteilung. Das Sprachsegment wird auf Basis der Transformationskoeffizienten in Klassen eingeteilt. Um eine ausreichend feine Zeitlauflösung zu erreichen, wird das Segment noch in P Subrahmen eingeteilt, so daß für jeden Subrahmen ein Klassifizierungsergebnis ausgegeben wird. Für einen Einsatz in niederratigen Sprachcodierverfahren wurde die Unterscheidung der folgenden Klassen vorgenommen:

(1) Hintergrundrauschen/stimmlos,
(2) Signalübergänge/"voicing onsets",
(3) Periodisch/stimmhaft.

Beim Einsatz in bestimmten Codierverfahren kann es sinnvoll sein, die periodische Klasse noch weiter aufzuteilen, etwa in Abschnitte mit überwiegend tieffrequenter Energie oder eher gleichmäßig verteilter Energie. Optional kann deshalb auch eine Unterscheidung von mehr als drei Klassen durchgeführt werden.

Im Anschluß daran erfolgt in einem entsprechenden Prozessor die Parameterberechnung. Zunächst wird aus den Transformationskoeffizienten $S_h(m,n)$ ein Satz von Parametern bestimmt, mit deren Hilfe dann anschließend die endgültige Klasseneinteilung vorgenommen werden kann. Die Auswahl der Parameter Skalierungs-Differenzmaß (P_1), zeitliches Differenzmaß (P_2) und Periodizitätsmaß (P_3) erwiesen sich dabei als besonders günstig, da sie einen direkten Bezug zu den definierten Klassen (1) bis (3) aufweisen.

— Für P_1 wird die Varianz der Energie der DWT-Transformationskoeffizienten über alle Skalierungsbereiche berechnet. Auf Basis dieses Parameters kann rahmenweise — also für ein relativ grobes Zeitraster — festgestellt werden, ob das Sprachsignal stimmlos ist bzw. nur Hintergrundrauschen vorliegt.
— Um P_2 zu ermitteln, wird zunächst die mittlere Energiedifferenz der Transformationskoeffizienten zwischen dem aktuellen und dem vergangen Rahmen berechnet. Nun werden für Transformationskoeffizienten feiner Skalierungsstufe (m klein) die Energiedifferenzen zwischen benachbarten Subrahmen ermittelt und mit der Energiedifferenz für den Gesamtrahmen verglichen. Dadurch kann ein Maß für die Wahrscheinlichkeit eines Signalübergangs (zum Beispiel stimmlos auf stimmhaft) für jeden Subrahmen — also für ein feines Zeitraster — bestimmt werden.
— Für P_3 werden rahmenweise die lokalen Maxima von Transformationskoeffizienten grober Skalie-

rungsstufe (m nahe bei M) bestimmt und geprüft, ob diese in regelmäßigen Abständen auftreten. Als lokale Maxima werden dabei die Spitzen bezeichnet, die einen gewissen Prozentsatz T des globalen Maximums des Rahmens übersteigen.

Die für diese Parameterberechnungen erforderlichen Schwellwerte werden in Abhängigkeit vom aktuellen Pegel des Hintergrundgeräusches adaptiv gesteuert, wodurch die Robustheit des Verfahrens in gestörter Umgebung gesteigert wird.

Darauffolgend wird die Auswertung vorgenommen. Die drei Parameter werden der Auswerteeinheit in Form von "Wahrscheinlichkeiten" (auf den Wertebereich (0,1) abgebildete Größen) zugeführt. Die Auswerteeinheit selbst trifft das endgültige Klassifizierungsergebnis für jeden Subrahmen auf Basis eines Zustandsmodells. Dadurch wird das Gedächtnis der für vorangegangene Subrahmen getroffenen Entscheidungen berücksichtigt. Außerdem werden nicht sinnvolle Übergänge, wie zum Beispiel direkter Sprung von "stimmlos" auf "stimmhaft", verboten. Als Ergebnis wird schließlich pro Rahmen ein Vektor mit P Komponenten ausgegeben, der das Klassifizierungsergebnis für die P Subrahmen enthält.

In den **Fig.** 2a und 2b sind die Klassifizierungsergebnisse für das Sprachsegment " ... parcel, I'd like ... " einer englischen Sprecherin exemplarisch dargestellt. Dabei wurden die Sprachrahmen der Länge 20 ms in vier equidistante Subrahmen zu jeweils 5 ms eingeteilt. Die DWT wurde nur für dyadische Skalierungsschritte ermittelt und auf Basis von kubischen Spline-Wavelets mit Hilfe einer rekursiven Filterbank implementiert. Die drei Signalklassen werden mit 0,1,2 in der gleichen Reihenfolge wie oben bezeichnet. Für **Fig.** 2a wurde Telefonband-Sprache (200 Hz bis 3400 Hz) ohne Störung verwendet, während für **Fig.** 2b zusätzlich Fahrzeuggeräusche mit einem durchschnittlichen Signal-Rausch-Abstand von 10 dB überlagert wurden. Der Vergleich der beiden Abbildungen zeigt, daß das Klassifizierungsergebnis nahezu unabhängig vom Rauschpegel ist. Mit Ausnahme kleinerer Unterschiede, die für Anwendungen in der Sprachcodierung irrelevant sind, werden die perzeptuell wichtigen periodischen Abschnitte sowie deren Anfangs- und Endpunkte in beiden Fällen gut lokalisiert. Durch Auswertung einer großen Vielfalt unterschiedlichen Sprachmaterials ergab sich, daß der Klassifizierungsfehler deutlich unter 5% für Signal-Rausch-Abstände oberhalb 10 dB liegt.

Der Klassifizierer wurde zusätzlich für folgenden typischen Anwendungsfall getestet: Ein CELP-Codierverfahren arbeitet bei einer Rahmenlänge von 20 ms und teilt diesen Rahmen zur effizienten Anregungscodierung in vier Subrahmen à 5 ms ein. Für jeden Subrahmen soll entsprechend der drei oben genannten Signalklassen auf Basis des Klassifizierers eine angepaßte Kombination von Codebüchern verwendet werden. Es wurde für jede Klasse ein typisches Codebuch mit jeweils 9 Bit/Subrahmen zur Codierung der Anregung eingesetzt, wodurch sich eine Bitrate von lediglich 1800 Bit/s für die Anregungscodierung (ohne Gain) ergab. Es wurden für die stimmlose Klasse ein Gauß'sches Codebuch, für die Onset-Klasse ein Zwei-Puls-Codebuch und für die periodische Klasse ein adaptives Codebuch verwendet. Schon für diese einfache, mit festen Subrahmenlängen arbeitende Konstellation von Codebüchern ergab sich eine gut verständliche Sprachqualität, jedoch noch mit rauhem Klang in periodischen Abschnitten. Zum Vergleich sei erwähnt, daß in ITU-T, Study Group 15 Contribution- Q. 12/15: Draft Recommendation G. 729 — Coding Of Speech at 8 kbit/s using Conjugate-Structure-Algebraic-Code-Excited-Linear-Predictive (CS-ACELP) Coding, 1995, für die Codierung der Anregung (ohne Gain) 4800 Bit/s benötigt werden, um Leitungsqualität zu erzielen. Selbst in Gerson, I. et al, Speech and Channel Coding for the Half-Rate GSM Channel, ITG-Fachbericht "Codierung für Quelle, Kanal und Übertragung", 1994, werden dafür noch 2800 bit/s verwendet, um Mobilfunkqualität sicherzustellen.

Patentansprüche

1. Verfahren zur Klassifizierung von Sprache, insbesondere Sprachsignalen für die signalangepaßte Steuerung von Sprachcodierverfahren zur Senkung der Bitrate bei gleichbleibender Sprachqualität oder zur Erhöhung der Qualität bei gleicher Bitrate, **dadurch gekennzeichnet**, daß nach einer Segmentierung des Sprachsignals für jeden gebildeten Rahmen eine Wavelet-Transformation berechnet wird, aus der mit Hilfe adaptiver Schwellen ein Satz Parameter (P_1 — $_3$) ermittelt wird, die ein Zustandsmodell steuern, das den Sprachrahmen in Unterrahmen aufteilt und jeden dieser Unterrahmen in eine von mehreren, für die Sprachcodierung typische Klassen unterteilt.
2. Verfahren nach Patentanspruch 1, dadurch gekennzeichnet, daß das Sprachsignal in Segmente konstanter Länge eingeteilt wird, und daß zur Vermeidung von Randeffekten bei der sich anschließenden Wavelet-Transformation entweder das Segment an den Grenzen gespiegelt wird, oder die Wavelet-Transformation im kleineren Intervall (L/2, N — L/2) berechnet wird und der Rahmen nur um den konstanten Versatz L/2 verschoben wird, so daß die Segmente sich überlappen oder daß an den Rändern des Segments mit den vorangegangenen bzw. zukünftigen Abtastwerten aufgefüllt wird.
3. Verfahren nach Patentanspruch 1 oder 2, dadurch gekennzeichnet, daß für ein Segment s(k) eine zeitdiskrete Wavelet-Transformation (DWT) $S_h(mn)$ bezüglich eines Wavelets h(k) mit den ganzzahligen Parametern Skalierung in und Zeitverschiebung n berechnet wird, und daß das Segment auf Basis der Transformationskoeffizienten in Klassen eingeteilt wird, insbesondere zur Erreichung einer feinen Zeitauflösung noch in P Subrahmen eingeteilt und für jeden Subrahmen ein Klassifizierungsergebnis errechnet und ausgegeben wird.
4. Verfahren nach einem der Patentansprüche 1 — 3, dadurch gekennzeichnet, daß aus dem Transformationskoeffizienten $S_h(mn)$ ein Satz von Parametern, insbesondere Skalierungs-Differenzmaß (P_1), zeitliches Differenzmaß (P_2) und Periodizitätsmaß (P_3) bestimmt wird, mit deren Hilfe dann anschließend die endgültige Klasseneinteilung vorgenommen wird, wobei die für diese Parameterberechnungen erforderlichen Schwellwerte in Abhängigkeit vom aktuellen Pegel des Hintergrundgeräusches adaptiv gesteuert werden.
5. Anordnung, insbesondere Klassifizierer zur Durchführung des Verfahrens nach einem der Patentansprüche 1 — 4, dadurch gekennzeichnet, daß die Eingangssprache einer Segmentierungseinrichtung zugeführt wird, daß nach der Segmentierung der Eingangssprache für jeden gebildeten Rahmen

bzw. für jedes gebildete Segment durch einen Prozessor eine diskrete Wavelet-Transformation berechnet wird, daß daraus mit Hilfe adaptiver Schwellen ein Satz Parameter (P_1 — P_3) ermittelt wird, die als Eingangsgrößen einem Zustandsmodell zugeführt werden, das seinerseits den Sprachrahmen in Unterrahmen aufteilt und jeden dieser Unterrahmen in eine von mehreren für die Sprachcodierung typische Klassen einteilt.

Hierzu 3 Seite(n) Zeichnungen

Nummer: DE 195 38 852 A1
Int. Cl.[6]: G 10 L 7/02
Offenlegungstag: 2. Januar 1997

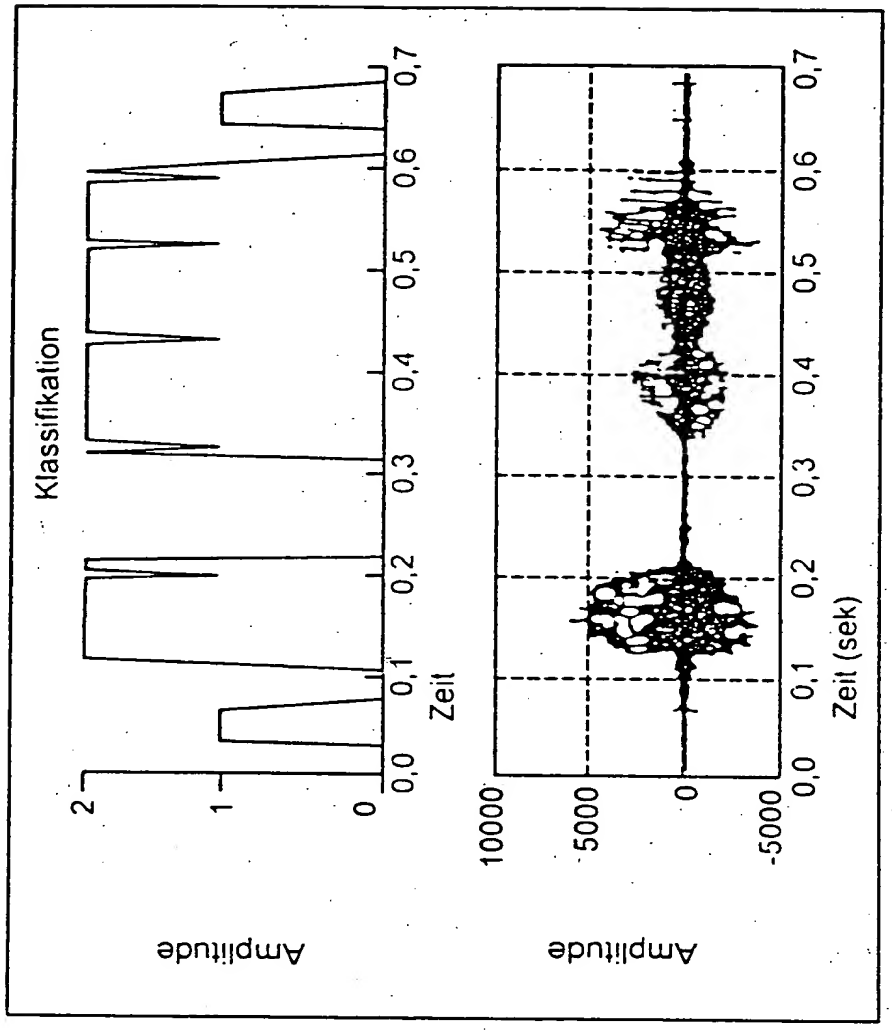


Fig. 2a

Nummer: DE 195 38 852 A1
Int. Cl.[6]: G 10 L 7/02
Offenlegungstag: 2. Januar 1997

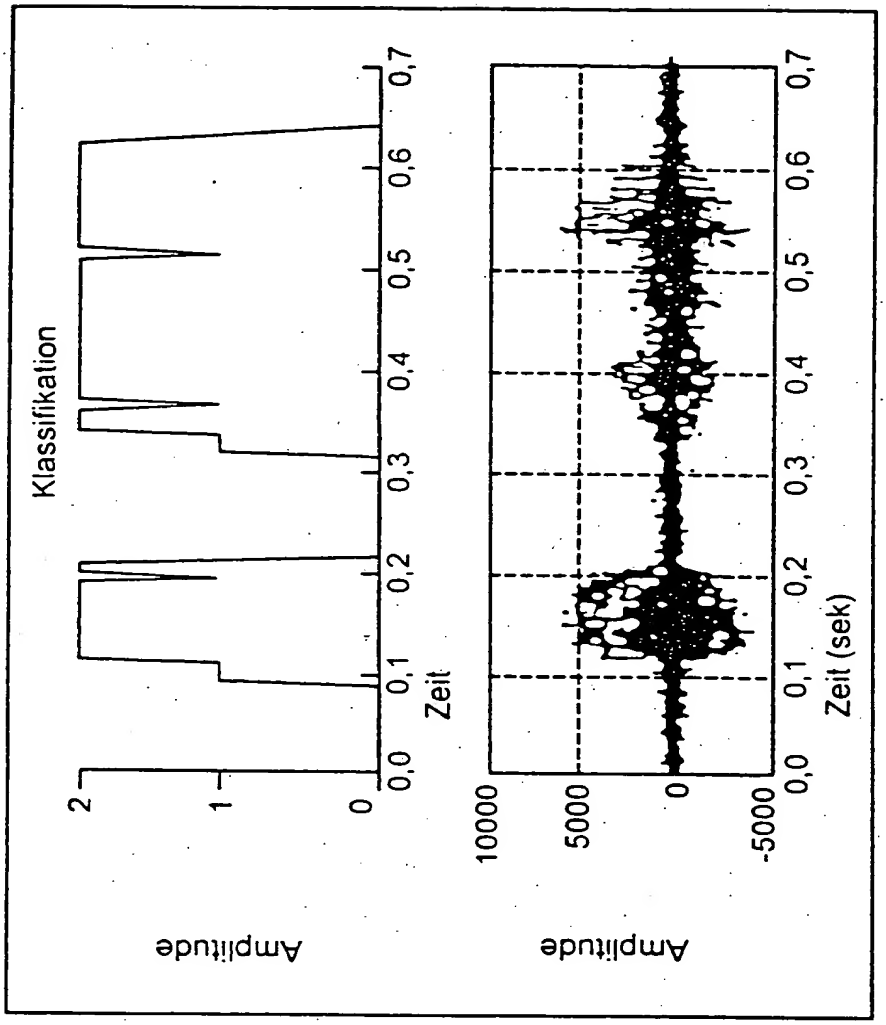


Fig. 2b

Nummer: DE 195 38 852 A1
Int. Cl.[6]: G 10 L 7/02
Offenlegungstag: 2. Januar 1997

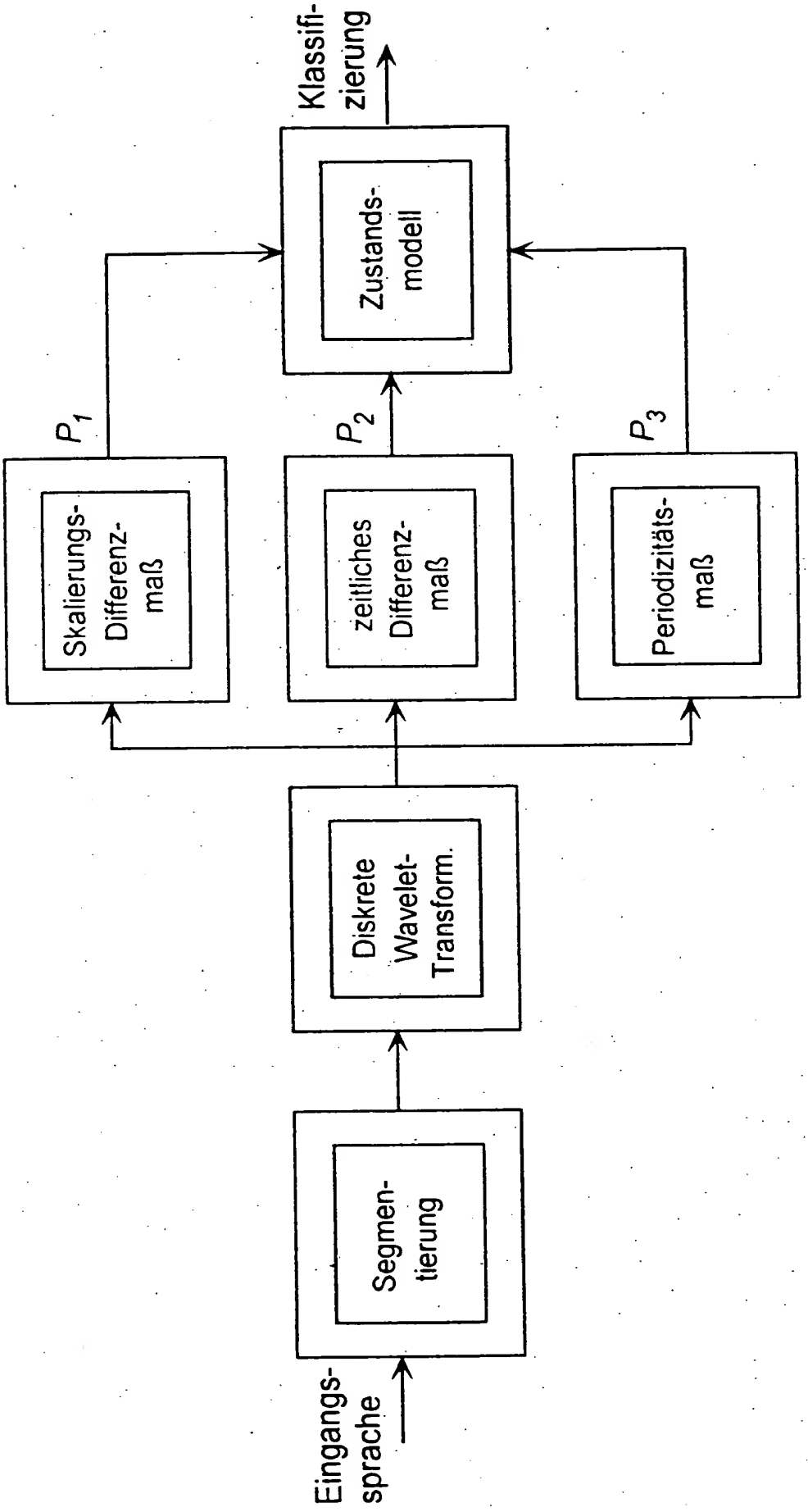


Fig. 1